AIQ: Reliability and Validity Evidence

The reliability of a test refers to the accuracy, consistency, and stability of test scores (Anastasi & Urbina, 1997). Current reliability studies for the AIQ include measures of internal consistency and test-retest reliability. The validity of a test refers to the degree to which the test measures what it is intended to assess. The validity of the AIQ is based on evidence from multiple sources, including (1) the content of the test items; (2) the intercorrelations among the AIQ subtests and composites; (3) the relationships between the AIQ and other measures (convergent and discriminant validity); and (4) the relationships between AIQ scores and outcomes of interest in sport (criterion-related validity).

Reliability Evidence:

Two sources of reliability are important for an intellectual ability measure, such as the AIQ. They are internal consistency and test-retest reliability (McGrew & Flanagan, 1998). The original normative set (N =299) was examined for the internal consistency analyses, while a subgroup (N = 48) was examined to determine test-retest reliability.

Internal Consistency

Internal consistency reflects the degree to which the items on a particular subtest are measuring the same underlying construct. The typical internal consistency statistic is Cronbach's alpha, which is the mean correlation of all possible sets of scores within a subtest. This statistic was judged to be appropriate to estimate the reliability for the subtests in which every athlete saw all items (viz., untimed tasks). Thus, estimates of reliability for timed subtests were determined via test-retest analyses. For the normative sample, internal consistency was moderate to high, with alpha coefficients ranging from .77 to .91.

Test-Retest Reliability

Test-retest reliability indicates the stability of a measure over time for abilities or traits that are presumed to remain relatively constant (i.e., intellectual ability). Test-retest reliability was examined for a normative subsample (N = 48). This form of reliability was conducted for the subtests that involved a time limit, wherein the examinees would see different numbers of items depending on how quickly they responded. The test-retest correlations across the subtests ranged from .57 to .84 (3-6 month interval), with most reliability coefficients considered to be good. Of note is that test-retest reliability coefficients for heavily speed tasks tend to be lower-bound estimates of a test's true reliability (McGrew & Flanagan, 1998).

Validity Evidence:

Evidence supporting the valid interpretation of given scores is based on various sources. In this case, the interpretation of the AIQ would be considered valid based on (a) the content of its items, (b) the intercorrelations among subtest and composite scores, (c) the convergence and divergence of AIQ scores with other measures (i.e., convergent and discriminant validity), and (d) the concurrent or predictive relationships between AIQ scores and outcomes of interest (i.e., criterion-related validity).

Content Validity

The relationship between a test's content and the construct it is intended to measure is a significant source of evidence about the validity of the test itself. This form of validity is not based on statistical analyses or empirical testing. Instead, it is based on the degree to which the test items adequately represent and capture the ability being measured. Test content also involves the format of items and the procedures for administering and scoring the test. An important goal in the development of the AIQ was to ensure that the items and subtests adequately sampled the specific aspects of athletic intelligence that the test is intended to measure. In particular, the items and subtests assess a range of cognitive abilities, including visual-spatial processing, long-term storage and retrieval, processing speed, and reaction time. As described in the AIQ Professional Manual, comprehensive literature reviews were conducted, and experts in the field were consulted in the development of the AIQ. In addition, Chapter 1 of the AIQ Professional Manual provides a detailed description of the theoretical rationale that guided item and subtest development.

Intercorrelation Analyses

In line with Cattell-Horn-Carroll Theory, two *a priori* hypotheses were made regarding the intercorrelations that would be found among the subtests and composites of the AIQ. First, because all subtests were designed to measure CHC cognitive abilities, it was assumed that there would be significant intercorrelations among all of the subtests and composite scores. In fact, each subtest was significantly correlated with every other subtest, with the exception of the reaction time subtests. The reaction time standard scores for Simple Reaction Time and Choice Reaction Time were not found to correlate with certain subtests. Although these subtests measure cognitive abilities outlined in CHC Theory, they also assess abilities that incorporate psycho-motor speed much more than the other subtests. As such, this finding is not surprising.

The second *a priori* hypothesis was that the subtests that are subsumed by a specific composite (e.g., visual spatial processing, long-term storage and retrieval, reaction time and processing speed) would have stronger correlations with each other than with the subtests that comprise other scales. For instance, the correlation between the Shape Rotations and Block Design subtests (two measures of visual spatial processing) would be expected to have stronger correlations than between either of these subtests and any of the long-term storage and retrieval, reaction time, or processing speed subtests (e.g., Paired Associative Learning, Simple Reaction Time, or Number Matching). In fact, all subtests were found to correlate the strongest with other subtests from the same composite, rather than subtests from other composites. The only exception to this finding is the Memory for Shapes subtest, a measure of visual memory. According to current research in CHC Theory (Flanagan, et al., 2013), the narrow ability of visual memory may be considered a visual-spatial processing ability; however, it has also been found to correlate strongly with measures of memory. As such, the current findings are generally consistent with *a priori* hypotheses and previous research.

Convergent and Discriminant Validity

One method of evaluating the validity of an instrument is to examine correlations between the test and other existing measures. Evidence of validity can be determined based on patterns of correlations, wherein higher correlations are found between the AIQ and other measures with which it theoretically should correlate, and lower correlations between the AIQ and measures with which it should not correlate.

To investigate this, a study was recently conducted in which athletes' scores on the AIQ were compared to obtained scores on the Wonderlic Personnel Test and the ImPACT test (Lovell, Collins, & Podell, 2000). The participants in this study included 93 Division 1 NCAA Men's Lacrosse, Men's Soccer, and Women's Soccer players attending a northeast university (Crimarco, O'Brien, & Bowman, 2020). Significant correlations were found between the Visual Spatial Processing and Long-term Memory factors of the AIQ and the Wonderlic test. However, neither the Reaction Time nor the Processing Speed factors of the AIQ correlated significantly with the Wonderlic test, thereby demonstrating discriminant validity. Significant correlations were also found in expected directions between composites of the ImPACT and factors on the AIQ. For instance, the AIQ Reaction Time and Processing Speed factors correlated significantly with the ImPACT Reaction Time composite, thereby demonstrating convergent validity. Overall, the results of this investigation provide further evidence for the construct validity of the AIQ.

Criterion-Related Validity

Another method to evaluate the validity of an instrument is to examine the relationships between test scores and concurrent or predictive data. Evidence of validity can be determined based on significant correlations or regression analyses involving AIQ scores and specific outcomes of interest in sports.

In the first of several studies, validity evidence for the AIQ was examined vis-à-vis outcomes in professional baseball (Bowman, Boone, Goldman, & Auerbach, 2020). Specifically, AIQ scores were obtained from 150 Minor League Baseball (MiLB) players prior to the start of the 2014 baseball season and their performance was then assessed through both traditional and new baseball statistics. It was hypothesized that the AIQ would demonstrate statistically significant relationships with outcomes of interest in both hitting and pitching categories, in hierarchical multiple regression analyses, after controlling for other variables, such as age, country of origin, and position. In fact, several significant relationships emerged between the AIQ and hitting performance. Additionally, an interaction emerged between visual spatial processing and reaction time that accounted for a statistically significant increase in the explanation of earned run average (ERA) for pitchers (Bowman et al., 2020). Overall, these findings suggest that players with higher AIQ scores tended to achieve better hitting and pitching outcomes.

To provide further evidence for the criterion-related validity of the Athletic Intelligence Quotient (AIQ), research was undertaken to examine the relationships among AIQ factors and performance outcomes in the National Football League (Bowman, Boone, Zaichkowsky, Goldman, & Auerbach, 2020). The results of this study revealed that specific AIQ factors accounted for a statistically significant increase in the explanation of variance in game statistics (e.g., rushing yards per carry) as well as overall ratings of player success (i.e., career approximate value) beyond other important factors (i.e., draft order).